A NOTE ON EFFECTIVE POPULATION SIZE WITH OVERLAPPING GENERATIONS[1]

WILLIAM G. HILL[2]

Department of Statistics, Iowa State University,
Ames, Iowa 50011

ABSTRACT

A simple derivation is given for a formula obtained previously for the effective size of random-mating populations with overlapping generations. The effective population size is the same as that for a population with discrete generations having the same variance of lifetime family size and the same number of individuals entering the population per generation.

IN recent years, formulae for rates of inbreeding and increment in genetic drift have been developed for populations with overlapping generations. FELSEN-STEIN (1971) described a haploid model with fixed numbers in each age class and parental age distribution. Subsequently, the model has been extended to random-mating diploid populations (HILL 1972a; JOHNSON 1977; CHOY and WEIR 1978; EMIGH and POLLAK 1979; see also these papers for a more complete bibliography). If generations overlap, the levels of inbreeding and drift variance among conceptual replicates do not increase at a steady rate, in contrast to the case of discrete generations; but eventually the increase does asymptote at a steady rate relative to the variability then present in the population. FELSENSTEIN (1971), JOHNSON (1977), CHOY and WEIR (1978) and EMIGH and POLLAK (1979) give exact recurrence formulae for probabilities of identity by descent, and JOHNSON (1977) also gives approximate formulae in a more elegant form. FELSENSTEIN, JOHNSON and EMIGH and POLLAK show how to compute the asymptotic rate of increase in inbreeding coefficient and the effective population size in terms of the members of individuals in each age class and the parental age distribution. HILL (1972a) showed that the effective sizes of populations with overlapping generations "equal the effective sizes of populations with discrete generations which have the same number of individuals entering the population each *generation* and the same variance of *lifetime* family member". For specific cases, this result has been checked by JOHNSON (1977) and by CHOY and WEIR (1978). The proof in my paper (HILL 1972a) is very tortuous, especially in view of the simple result. A much easier and perhaps more enlightening derivation follows.

ANALYSIS

To clarify the presentation, let us consider first a haploid model, then a monoecious diploid model and finally a dioecious diploid model. In each case there are assumed to be discrete breeding periods, e.g., years, a constant number born or chosen for breeding at each time period, and a constant parental age distribution such that the generation interval (mean age of parents at birth of their progeny) remains fixed. The effects are studied of sampling genes that are neutral with respect to fitness among the individuals present some period a long time in the past on the variation in mean gene frequency among conceptual replicate populations now. It is essential that the intervening time be long, so that the genes present in that cohort have had a chance to spread through all age groups.

Haploid model: Assume that N individuals are born each time period and that the generation interval (mean age of parents when their progeny are born) is L. At some time in the distant past, assume that the cohort of N newborns had frequencies $X_1, \ldots, X_N$, where $X_i = 0$ or 1, of some neutral gene, with $\sum_{i=1}^{N} X_i/N = q$ being the mean gene frequency. Assume also that individual i had n_{ij} progeny when aged j time units, and $n_i = \sum_{j=1}^{k} n_{ij}$ progeny in its whole life, where k is the age after which all reproduction ceases. For any newborn gene or individual (in the haploid case), the probability that it ultimately remains in the population is given by EMIGH (1979) and is equal to $1/(LN)$, which is the expected genetic contribution by the whole cohort $(1/L)$ divided by the number in the cohort (N). This probability does not depend on when it was born as long as we look far enough ahead subsequently. An alternative viewpoint is that the contribution of the newborn to the subsequent mean gene frequency is a fraction $1/L$ of its gene frequency (0 or 1). Let the mean contribution by the whole cohort be Q, say, where

$$Q = \frac{1}{LN} \sum_{i=1}^{N} n_i X_i \ . \tag{1}$$

Differential reproduction among the individuals will induce a change in Q and cause genetic drift. The form of (1) is, however, just the same as that for a cohort of individuals with discrete generations, except for the scale factor, $1/L$. Thus, methodology and results of LATTER (1959) and of CROW and KIMURA (1970) can be used directly. The mean of the X_i is q and the mean family size, averaged over all cohorts, is $\mu_n = 1$, since the population is of constant size. Thus, the expected value of Q is q/L, and the genetic drift is, from (1),

$$Q - q/L = \frac{1}{LN} \sum_{i=1}^{N} (n_i X_i - q)$$

$$= \frac{1}{LN} \sum_{i=1}^{N} (n_i - \mu_n) X_i \ .$$

Let ue define the variance of family size as $\sigma_n{}^2 = \sum_{i=1}^{N} (n_i - \mu_n)^2 / N$ and, for simplicity, ignore terms in $1/N^2$ relative to $1/N$ that are introduced by correlations among the n_i and among the X_i since their sums are fixed. The n_i and X_i are independent and $V(X_i) = q(1-q)$, hence

$$V(Q-q/L) = q(1-q) \ \sigma_n{}^2 / (NL^2) \ . \tag{2}$$

To compute the effective population size, consider an idealized population of N haploid individuals with no differential viability or fertility and discrete generations, where the variance in family size equals 1 from the Poisson distribution). More strictly, the variance equals $1 - 1/N$ from the multinomial, but second order terms have already been ignored in (2). The drift from one generation of sampling given a mean gene frequency of q, observed either in the next generation or in terms of ultimate fixation, is $q(1-q)/N$. The "annual effective population size", N_a, the effective size of a population having discrete generations of length one time period or, for example, one year and the same increment in drift variance (HILL 1972a) is thus, using (2),

$$\begin{aligned} N_a &= q(1-q)/V(Q-q/L) \\ &= NL^2/\sigma_n{}^2 \end{aligned} \tag{3}$$

The effective population size measures the increment in drift per generation, or per L time periods and is, assuming that the increments in drift are small so that they are essentially linear over one generation,

$$\begin{aligned} N_e &= N_a/L \tag{4} \\ &= NL/\sigma_n{}^2 \ . \tag{5} \end{aligned}$$

This is the same formula as derived previously (HILL 1972a), except that second order terms were included, to give $N_e = (N-1)L/\sigma_n{}^2$.

Monoecious diploids: Assuming random mating, but without any need to specify the joint age distribution of mates, the analysis extends readily to monecious diploids with random selfing. If N individuals are born in each cohort, the probability of ultimate survival of any gene is $1/(2NL)$ (EMIGH and POLLAK 1979), rather than $1/2N$ with discrete generations. Discrete-generation methods (CROW and KIMURA 1970) can now be used and scaled by the factor $1/L$, and are given here only for completeness. Drift in the mean gene frequency of progeny arises both from differential progeny numbers and from segregation among heterozygotes. As before, let X_i be the frequency of genes of one type in individual i in a cohort born a long time ago, with X_i now taking values of 0, 1/2 and 1, and let n_i be the number of progeny in its lifetime (for stable population size the mean number of progeny per individual equals two). Also let δ_{ij} be the difference in frequency between the jth sampled gene and its parental value X_i,

where $\delta_{ij} = 0$ if the parent is a homozygote or $\pm 1/2$ if a heterozygote. Then corresponding to (1),

$$Q = \frac{1}{2LN} \left[\sum_{i=1}^{N} (n_i X_i + \sum_{j=1}^{n_i} \delta_{ij}) \right] . \tag{6}$$

The components of (6) are sampled independently. Noting that $V(X_i) = q(1-q)/2$ and $V(\delta_{ij}) = E(\delta_{ij}^2) =$ frequency of heterozygotes times variance given heterozygotes $= 2q(1-q)(1/4) = q(1-q)/2$,

$$V(Q-q/L) = q(1-q)(\sigma_n^2 + 2)/(8NL^2), \tag{7}$$

again with terms of order $1/N^2$ ignored.

In the idealized population with discrete generations, the drift variance per generation is $q(1-q)/2N$. Hence, from (7) and (4),

$$N_a/L = N_e = 4NL/(\sigma_n^2 + 2) , \tag{8}$$

as given previously (HILL 1972a), apart from second order terms. With a Poisson distribution of family sizes, $\sigma_n^2 = 2$ and $N_e = NL$; but with random deaths and family sizes at each breeding period, variance of lifetime family size exceeds that from the Poisson (HILL 1972a).

Two sexes: The numbers used in the above calculations could equally be in terms of the numbers born in each cohort, in which case σ_n^2 includes differences in viability both before and after reaching reproductive age, or in terms of numbers reaching breeding age, in which case pre-reproductive viability differences are excluded. To allow for either approach in the dioecious case, assume that different numbers, M males and F females, are born or sampled for breeding in each cohort. There are now four pathways for genes:

Pathway		mean age	family size	Heterozygote sampling	Family size variance	covariance
Parent	Progeny					
male	– male	L_{mm}	n_{mmi}	δ_{mmij}	σ^2_{mmi}	cov(mm,mf)
male	– female	L_{mf}	n_{mfi}	δ_{mfij}	σ^2_{mfi}	
female	– male	L_{fm}	n_{fmi}	δ_{fmij}	σ^2_{fmi}	cov (fm,ff)
female	– female	L_{ff}	n_{ffi}	δ_{ffij}	σ^2_{ffi}	

The generation interval is given by $L = (L_{mm} + L_{mf} + L_{fm} + L_{ff})/4$, and the probability of ultimate survival of a gene is $1/(4ML)$ in a male of pre-reproductive age and $1/(4FL)$ in a female (EMIGH and POLLAK 1979). The contribution of a cohort to ultimate mean gene frequency is

$$Q = \frac{1}{4ML} \left[\sum_{i=1}^{M} (n_{mmi} X_{mi} + \sum_{j=1}^{n_{mmi}} \delta_{mmij}) + \sum_{i=1}^{F} (n_{fmi} X_{fi} + \sum_{j=1}^{n_{fmi}} \delta_{fmij}) \right]$$

$$+ \frac{1}{4FL} \left[\sum_{i=1}^{M} (n_{mfi} X_{mi} + \sum_{j=1}^{n_{mfi}} \delta_{mfij}) + \sum_{i=1}^{F} (n_{ffi} X_{fi} + \sum_{j=1}^{n_{ffi}} \delta_{ffij}) \right] .$$

The drift variance follows immediately, and by using discrete generation results,

$$\frac{L}{N_a} = \frac{1}{N_e} = \frac{1}{16ML} \left[2 + \sigma_{mm}^2 + 2\left(\frac{M}{F}\right)\text{cov}(mm,mf) + \left(\frac{M}{F}\right)^2 \sigma_{mf}^2\right]$$

$$+ \frac{1}{16FL} \left[2 + \left(\frac{F}{M}\right)^2 \sigma_{fm}^2 + 2\left(\frac{F}{M}\right)\text{cov}(fm,ff) + \sigma_{ff}^2\right] . \tag{9}$$

If $M = F = N/2$, for example, with counting in newborns

$$N_e = 8NL/(\sigma_m^2 + \sigma_f^2 + 4) , \tag{10}$$

where $\sigma_m^2 = \sigma_{mm}^2 + \sigma_{mf}^2 + 2\text{cov}(mm,mf)$ is the variance of the total number of progeny of a male, and σ_f^2 is defined similarly. If $\sigma_m^2 = \sigma_f^2$, (10) reduces to (8).

DISCUSSION

The results, (5), (8), (9) and (10), reiterate those obtained previously (HILL 1972a), differing only in the simplicity of their derivation, and confirming that the variance effective number with overlapping generations is the same as with discrete generations, providing lifetime family size and numbers entering each generation are used. The population numbers cannot be too small, or there will be wild fluctuations in gene frequency over a period of a generation and the drift per generation will not equal L times that per year. The formal need for a stable age distribution must also be emphasized, otherwise progeny born at different periods of an individual's life will have different chances of contribution subsequently. However, it seems quite reasonable to assume that small disturbances either in age distribution or in the numbers born in each cohort will not substantially affect the results. A formal proof of this conjecture would be difficult.

The calculations have been made solely for the variance effective number. A similar argument can be adopted using the contribution of one cohort to the subsequent increase in inbreeding to show that inbreeding effective numbers are the same, as they must be in random mating populations of constant size (CROW and KIMURA 1970).

The formulae for effective population size given here in terms of variances of family size appear quite different from those of FELSENSTEIN (1971), JOHNSON (1977) and EMIGH and POLLAK (1979) which are expressions in numbers in the population and contributions to total reproduction of each age group. Their formulae can, however, be rearranged in terms of variance of family size, and then agree with those given here (JOHNSON 1977). The formulae in terms of variances of family size are more general, for they allow there to be differential viability and fertility, and correlations of fertilities of individuals at successive ages. In practice, they are most likely to be useful in getting an estimate of rates of inbreeding and drift in populations in which full pedigrees are not available, or would be expensive to utilize completely, or in organizing population structure to minimize rates of inbreeding and gene frequency drift, for example in an unselected control population (HILL 1972b).

Finally, it should be emphasized that these effective population size formulae give approximations only to the asymptotic rates of increase in drift or levels of inbreeding. Initially inbreeding increments are erratic and the level, even when

the rate of increase has settled down, may depart by a constant value from the product of number of time periods and asymptotic rate of increase per time period; even so the approximations have been found to be quite good (JOHNSON 1977; CHOY and WEIR 1978). Thus for general planning of animal breeding programs or study of long standing populations in the wild, these asymptotic rates should be an adequate guide.

LITERATURE CITED

CHOY, S. C. and B. S. WEIR, 1978   Exact inbreeding coefficients in populations with overlapping generations. Genetics **89**: 591–614.

CROW, J. F. and M. KIMURA, 1970   An Introduction to Population Genetics Theory. Harper and Row, New York.

EMIGH, T. H., 1979   The dynamics of finite haploid populations with overlapping generations. 1. Moments, fixation probabilities, and stationary distributions. Genetics **92**: 323–337.

EMIGH, T. H. and E. POLLAK, 1979   Fixation probabilities and effective population numbers in diploid populations with overlapping generations. Theoret. Popul. Biol. **15**: (in press).

FELSENSTEIN, J., 1971   Inbreeding and variance effective numbers in populations with overlapping generations. Genetics **68**: 581–597.

HILL, W. G., 1972a   Effective size of populations with overlapping generations. Theoret. Popul. Biol. **3**: 278–289. ——, 1972b   Estimation of genetic change. 1. General theory and design of control populations. Anim. Breed. Abstr. **40**: 1–15.

JOHNSON, D. L., 1977   Inbreeding in populations with overlapping generations. Genetics **87**: 581–591.

LATTER, B. D. H., 1959   Genetic sampling in a random mating population of constant size and sex ratio. Aust. J. Biol. Sci. **12**: 500–505.